

Robert H. Frank

Altruists with Green Beards: Still Kicking?

Abstract: In earlier work, I proposed the ‘adaptive standard of rationality’, according to which narrow self-interest models can be broadened by positing additional tastes, but only upon a plausible showing that those tastes do not hamper resource acquisition in competitive environments. This proposal is related to the green beard hypothesis from biology, according to which altruism might be adaptive if its presence could be reliably signaled by some observable feature, such as a green beard. In their contribution to this issue Ernst Fehr and Urs Fischbacher offer theoretical arguments and describe laboratory experiments whose results they interpret as refuting my version of the green beard hypothesis. In this response, I argue that their theoretical arguments and experimental evidence pose no threat to the green beard hypothesis.

1. Introduction

If self-interest is not the only human motive, it is surely an important one. Only partly for the sake of simplicity, many social scientists favor behavioral models that assume people are driven exclusively by narrow self-interest. A strong point in favor of these models is that their predictions about observable behavior are often confirmed.

When speeding fines increase, for example, motorists drive more slowly. When fuel prices rise, people turn down their thermostats and buy more fuel-efficient cars. When diners tip generously, they are greeted more warmly on subsequent visits. And so on.

Despite these successes, self-interest models often fail to track the data. Although they predict universal defection in one-shot prisoner’s dilemmas, for example, cooperation in this game is common.¹ People tip at about the standard rate, even when dining in restaurants they will never visit again (Bodvarsson/Gibson 1994). They often donate anonymously to charity and return lost wallets with the cash intact (Frank 1988).

Some economists respond to such observations by incorporating a broader mix of human motives into their behavioral models—arguing, for instance, that people give to charity because they get a warm glow from doing so (see Andreoni 1995). In effect, these economists adopt what I will call the ‘warm-glow standard’ of rationality, which holds that an individual is rational if she is efficient in pursuit of whatever aims she may happen to hold.

Although many find its additional flexibility attractive, the warm-glow standard has proved controversial among both economists and biologists. George

¹ For a meta-analysis of experimental findings for the prisoner’s dilemma, see Sally 1995.

Stigler and Gary Becker objected that because it can explain virtually any behavior by simply positing a taste for it, the warm-glow model is not really a scientific theory (Becker/Stigler 1977). Biologists, for their part, have argued that evolutionary forces should not favor motives that lead organisms to incur avoidable costs (Dawkins 1976). On that view, natural selection should favor individuals who get a warm glow from *not* giving to charity.

In my 1988 book, *Passions Within Reason*, and in subsequent work (2004), I explored whether self-interest models could be broadened in ways that meet the objections to the warm-glow standard. In response to the methodological concern expressed by Becker and Stigler, I proposed the ‘adaptive standard of rationality’. Like the warm-glow standard, it permits the self-interest model to be broadened by positing additional tastes, but only upon a plausible showing that those tastes do not hamper resource acquisition in competitive environments (see also Hirshleifer 1987).

Although biologists might object that this requirement militates against the emergence of a warm glow for giving to charity, I argue that under certain conditions it does not. Specifically, I show that people known to be driven by unselfish motives may for that very reason be better equipped than others to solve prisoner’s dilemmas and various other commitment problems.

My proposal is related to an earlier idea from biology, according to which altruism might be adaptive if its presence could be reliably signaled by some observable feature. Because the specific feature used in early discussions was a green beard, this idea became known as the green beard hypothesis.

In their contribution to this issue, *Altruists with Green Beards*, Ernst Fehr and Urs Fischbacher offer theoretical arguments and describe laboratory experiments whose results they interpret as refuting my version of the green beard hypothesis. In this response, I make three points: 1) that Fehr and Fischbacher misread my hypothesis as requiring that individuals be able to make accurate inferences about behavioral predispositions about randomly chosen strangers on the basis of brief personal interaction; 2) that their experimental findings, which focus narrowly on that ability, therefore shed no light on my hypothesis; and 3) that they err, on both theoretical and empirical grounds, in claiming that mimicry always precludes strategic signaling. In sum, I will argue that their theoretical arguments and experimental evidence pose no threat to the green beard hypothesis.

2. A Simple Thought Experiment

To introduce the model Fehr and Fischbacher believe they have refuted, I organize my comments around the following thought experiment:

You are the owner of successful local business. After careful study, you conclude that an outlet of your business would thrive in a similar city located 200 miles away. You cannot manage the outlet yourself, and the limitations of external auditing and control mechanisms will prevent you from knowing whether a hired manager has cheated you.

If you could hire an honest manager, you could pay him \$100,000 (twice the going rate) and still expect a net gain of \$100,000 by opening the outlet. The difficulty is that any manager you hire will face powerful incentives to cheat. By managing dishonestly, he could augment his own return by \$40,000, in which case you would lose \$100,000 on the venture. In that event, you would have no recourse, since there would be no way to know, much less prove, that your manager had cheated you. Would you open the branch outlet?

Before you respond, let us review how conventional neoclassical economic analysis would portray this decision. The information given in the thought experiment is summarized in the decision tree shown in Figure 1. At point A, you must decide whether to open the outlet. If you do not (that is, if you take the bottom branch of the decision tree at A), the scenario ends at C. You neither gain nor lose any money, and the managerial candidate you failed to hire earns the going rate of \$50,000 in alternative employment. If you choose instead to open the outlet (taking the top branch of the decision tree at A), the manager you hire must then decide at B whether to manage honestly. If so (taking the top branch at B), the scenario ends at D. You and the manager both earn \$100,000 from the venture. But if the manager takes the bottom branch at B, the scenario ends at E. By managing dishonestly, he earns \$140,000 while you suffer a loss of \$100,000.

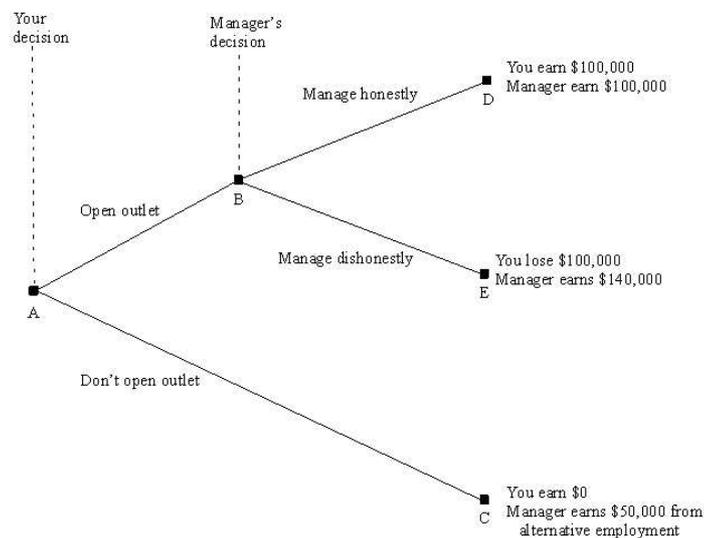


Figure 1: Decision Tree for the Outlet Problem

Following the conventions of economic theory, assume that you wish to open the outlet in the distant city only if you can reasonably expect it to make money

by doing so. Your task is thus to predict how the manager you hire would evaluate the choices confronting him at B. Economic theory conventionally assumes narrow self-interest on the part of managers and other hired agents. Such a manager would choose the bottom branch at B, since that would net him a payoff higher by \$40,000 than the amount he would receive on the top branch. And because that choice results in a loss of \$100,000 for you, your best bet would be not to open the outlet. After all, your payoff at C on the bottom branch from A (\$0) is better than your predicted payoff if you choose the top at A and hire a self-interested manager (-\$100,000).

As just described, the outlet decision problem constitutes what Tom Schelling (1960) called a ‘commitment problem’. If you assume that potential managers are self-interested in the narrow sense, you will not open the outlet, and this is a worse outcome for both you and the manager than if you had opened the outlet and the manager had run it honestly. If a managerial candidate could somehow commit himself to behave honestly he would clearly want to do so.

My essential claim in *Passion Within Reason* was that moral sentiments like the ones described by Adam Smith and David Hume could be favored by natural selection, provided that certain conditions were met. In the context of the decision problem just described, the key requirement is that owners be able to identify a candidate who would manage honestly with at least a statistical threshold level of reliability. Suppose, for example, that as the owner confronting the outlet decision problem, a managerial candidate whom you believe to be honest is in fact honest with probability 0.7. Referring to the payoffs summarized in Figure 1, we then see that if you open the outlet, your expected payoff will be $(0.7)(\$100,000) + (0.3)(-\$100,000) = \$40,000$. In this example, if owners could distinguish honest individuals from dishonest ones with 70 percent accuracy, they would open branch outlets that would be profitable 70 percent of the time. An honest managerial candidate’s expected payoff would be $(0.7)(\$100,000) + (0.3)(\$50,000) = \$85,000$, while a dishonest candidate’s expected payoff would be $(0.7)(\$50,000) + (0.3)(\$140,000) = \$77,000$. So it is clear that if sufficiently accurate predictions of honest behavior could be made, someone who was predisposed not to cheat in such circumstances could survive the pressures of natural selection.

Can you think of a specific individual (not related to you by blood or marriage) who you feel 70 percent sure would manage your outlet honestly if he or she happened to be available for the assignment? Faced with this question, most people say yes. If you agree, then you accept the central premise of the adaptive rationality standard I outlined in *Passions Within Reason*. Situations that constitute one-shot social dilemmas in the domain of material payoffs can sometimes be resolved if we are able to predict non-opportunistic behavior with sufficient precision.

The account just described does not require us to abandon rational choice theory. If you are sufficiently confident that a particular manager will not cheat you, you are rational to open the branch outlet. And if your prediction about the manager’s character is correct, he will rationally choose not to cheat you once he is hired. What my account adds to the conventional self-interest story is that

someone who was motivated (e.g. by a sympathetic bond) not to cheat his trading partner even though there would be no chance of being caught and punished might still do reasonably well in terms of his material payoffs. A willingness to exercise moral restraint even in one-shot dilemmas can thus prove adaptive under certain circumstances. Conversely, a person's single-minded pursuit of self-interest can prove maladaptive (provided others can identify and selectively avoid that person in ventures that require trust).

3. The Fehr-Fischbacher Experiments

Although Fehr and Fischbacher embed their discussion in the context of a somewhat different commitment problem (namely, the one-shot prisoner's dilemma), they are accurate in their description of the basic claim I illustrate here in the branch outlet decision. But they go on to report the results of experiments they interpret as implying that owners could not resolve the dilemma in the manner described. It is this interpretation that proves problematic.²

What their experiments show is that subjects who interact with one another only briefly can predict, with significantly better than chance accuracy, whether their partners will cooperate, but that their accuracy levels fall short of the threshold necessary to make cooperation a winning strategy. In the context of the branch outlet decision, their claim is equivalent to saying that owners can identify honest managers with better than chance accuracy, but not accurately enough to make the branch outlet a winner. But the question posed in the original thought experiment was not whether you could accurately predict whether someone of extremely limited acquaintance would manage honestly. Indeed, had that been the question, most people would have said no. Rather, the question was whether you could identify *someone* who would. And to that question most people respond affirmatively.

Typically, the persons they name are friends of long duration. Such choices seem natural for two reasons. First, the more time one spends with someone else, the more opportunities there are to observe clues to that person's emotional makeup. And second, the more time people spend with a friend, the deeper their emotional bonds are likely to be. Sympathy, affection, and other emotions that motivate trustworthy behavior are likely to be more strongly summoned by interactions with close friends than with perfect strangers.

Notice that although the people named are usually ones with whom we engage in repeated interactions, the particular episode involving the branch outlet is not a repeated game. After all, dishonest management is only one among multiple reasons a branch outlet might fail. So managing dishonestly would not lead to retaliation in the future.

When asked, most people say they named the people they did because they felt they knew them well enough to be able to say that they would *want* to

² I use the branch outlet problem as my illustration here because its simplicity helps highlight the basic flaw in the authors' critique of the green beard hypothesis.

manage the branch outlet honestly. In effect, the prospect of cheating a close friend would make them feel so bad that it just wouldn't be worth it.

My account of the adaptive rationality standard does not require that persons who are predicted to behave honestly be irrevocably bound to do so by any specific emotion. A person might find the idea of cheating his employer morally repugnant, yet succumb to temptation anyway. Nor does my account require that behavioral predispositions be unconditional. Smith and Hume believed, as I do, that cooperation is often the result of bonds of mutual sympathy between exchange partners. In such cases, any behavioral predisposition to cooperate is necessarily a highly complex and conditional one. Certainly the evidence supplies no reason to suppose that cooperation is an unconditional behavioral trait. In my experiments with Gilovich and Regan (1993), for example, many subjects cooperated with one of their partners and defected with the other.

What the adaptive rationality standard requires is that there be cues that lead to correct, situation-specific predictions with sufficient frequency. Typically, the persons about whom predictions of cooperation are made with confidence are drawn from a small subset of close friends of long duration. In contrast, the experiments performed by Fehr and Fischbacher were performed with subjects who were strangers at the outset and interacted with one another only briefly. These experiments simply do not speak to the question of whether it is possible to identify someone who would manage a branch outlet honestly.

4. Mimicry and Strategic Signaling

Fehr and Fischbacher also argue that strategic signaling of the type I describe should not be possible on theoretical grounds:

“... the crucial assumption behind this approach is that there are no selfish mutants with green beards. As soon as one allows for such mutants, the argument breaks down because the mutants reap the same benefits as the altruists but do not bear the cost of altruistic acts.” (Fehr/Fischbacher, *Green Beards*, this issue, 74)

Fehr and Fischbacher are correct to join others in pointing out that defectors have an obvious incentive to mimic whatever signs we use for identifying reliable trading partners. Selection pressure should therefore favor capacities for deception, and examples of such capacities clearly abound in human interaction. If signals of emotional commitment could be mimicked perfectly and without cost, these signals would eventually cease to be useful. Over time, natural selection would mould false signals into perfect replicas of real ones, driving the capacity for signaling commitment to extinction.

But the mere observation that costless, perfect mimicry would render a signal useless does rule out the possibility of an equilibrium that entails strategic signaling. Natural selection might be good at building a copy of a useful signal, but it might also be good at modifying an existing signal to evade mimicry. And because the original signal often has a substantial head start in this process, it

may be a difficult target. At the very least it is a moving one. Whether attempts to mimic it could keep pace simply cannot be settled on *a priori* grounds.

The signal in question might also be one that others simply cannot mimic. Long, brightly colored tail feathers, for example, are thought to signal robust health in peacocks, and hence to explain why peahens exhibit a strong preference for males with such displays. Why don't males in poor health just mimic the colorful displays of their rivals? Because they cannot. The observed preference of peahens is thought to be a stable equilibrium because unhealthy males simply cannot muster the physiological resources needed to support bright displays.

Theoretical considerations from the animal signaling literature identify another important difficulty confronting mimicry. These considerations suggest that because signals typically originate for reasons unrelated to their eventual function as conveyors of strategic information, they are often insulated against mimicry for extended periods. By the time they emerge as functional signals, they are often complex and highly idiosyncratic, making them far more difficult to mimic.

The basic problem for the emergence of strategic signals is that natural selection cannot be forward-looking. It cannot recognize, for example, that a series of mutations might eventually produce an individual able to signal its capacity to solve one-shot social dilemmas, and then favor the first costly step to that end, even though it yields no immediate benefit. It is this first step that presents the difficulty, because the initial appearance of a signal would have no meaning to external observers. It would thus entail costs, but no benefits. And the Darwinian rule is that a mutation must offer an *immediate* surplus of benefits over costs, or else be consigned to the evolutionary scrap heap.

How do signals ever originate, then? Essentially by accident, according to the derivation principle developed by Niko Tinbergen (1952). The constraint imposed by this principle is clearly illustrated by the example of the dung beetle. The insect gets its name from the fact that it escapes from predators by virtue of its resemblance to a fragment of dung. Biologists argue, however, that this advantage cannot explain how this beetle came to resemble a fragment of dung in the first place. The problem is that if we start with a species whose individuals bear not the slightest resemblance to a dung fragment, a minor mutation in the direction of a dung-like appearance would not have been of any use, since, as Stephen Jay Gould asks, "... can there be any edge in looking 5 percent like a turd?" (Gould 1977, 104) A mutation toward dung-like appearance will enhance fitness only if the individual's appearance *already* happened to be similar enough to a dung fragment for the mutation to have fooled the most myopic potential predator.

Thus the initial path toward near-resemblance must have been essentially a matter of chance—the result of mutations that were favored for other reasons and just happened to produce a dung-like appearance in the process. Once the resemblance crosses the recognition threshold by chance, however, natural selection can be expected to fine-tune the resemblance, in the same ruthlessly effective way it fine-tunes other useful traits.

Essentially the same logic should apply to the emergence of an observable

signal of a moral emotion such as sympathy. If the *only* behavioral effect of having sympathy were to motivate cooperation in one-shot prisoner's dilemmas, the first mutants with a small measure of this emotion would have enjoyed no advantage, even if their mutation happened to be accompanied by an observable signal. By virtue of its novelty, no one would have known what the signal meant, so it could not have facilitated selective interaction among sympathetic individuals. And since an indiscriminating tendency to cooperate entails costs, natural selection should have worked against sympathy, for the reasons just described.

If sympathy and other moral emotions were favored by natural selection in their earliest stages, they must therefore have conferred some other benefit. For example, perhaps a mutant with the capacity for sympathy was a more effective parent, a fitness enhancement that might have compensated for the initial costs of an indiscriminately sympathetic posture toward unrelated individuals. But because people are able to discriminate between kin and non-kin, it seems unlikely that sympathy selected for just this purpose could sustain cooperation with unrelated partners in one-shot social dilemmas.

A more promising possibility is that emotions like sympathy may function as self-control devices. In a world populated by utility maximizers of the sort usually assumed in economics, self-control problems would not exist. Such individuals would discount future costs and benefits at a constant exponential rate, which means that any choice that would seem best right now would also seem best in hindsight. Extensive evidence summarized by George Ainslie (1992), however, suggests that all creatures, animal and human, tend to discount future rewards not exponentially but hyperbolically. As Ainslie explains, hyperbolic discounting implies a temporary preference for 'the poorer but earlier of two goals, when the earlier goal is close at hand'. Seated before a bowl of salted cashews, for example, people often eat too many, and then later express sincere regret at having spoiled their dinners.

A similar time-inconsistency problem confronts people who interact in a sequence of repeated prisoner's dilemmas. In such situations, Rapport and Chammah, Axelrod, and others have demonstrated the remarkable effectiveness of the tit-for-tat strategy—in which you cooperate in the first interaction, then in each successive interaction mimic whatever your partner did in the immediately preceding one (Rapport/Chammah 1965; Axelrod 1984). Note, however, that implementation of tit-for-tat entails an inherent self-control problem. By cooperating in the current round, the tit-for-tat player must incur a present cost in order to receive a potentially much larger benefit in the future. In contrast, a player who defects in the current round receives a benefit immediately, whereas the costs of that action are both delayed and uncertain. Thus someone might realize he would come out ahead in the long run if he cooperated in the current interaction, yet find himself unable to resist the temptation to reap the immediate gains from defecting.

A person who is sympathetic toward potential trading partners is, by virtue of that concern, less likely than others to yield to temptation in the current interaction. Such a person would still find the gains from defecting attractive,

but their allure would be mitigated by the prospect of the immediate aversive psychological reaction that would be triggered by defecting. For this reason, persons with sympathy for their trading partners would find it easier than others to implement the tit-for-tat strategy in repeated prisoner's dilemmas. To the extent that the ability to execute tit-for-tat enhances fitness, people who experienced sympathy would have fared better than those who did not, even if no observable signal of sympathy were generally recognized.³

But given that external markers of sympathy exist, there is every reason to expect natural selection to have refined them for signaling purposes once they become recognized. We know, for example, that individual differences in emotional responsiveness are at least weakly heritable (Bruell 1970). If selective trustworthiness is advantageous and observable, natural selection should favor individual variants who are both more trustworthy and better able to communicate that fact to others.

For individuals who are not trustworthy, there would of course be advantage in being able to mimic the external markers of trustworthiness. But if an emotion was originally selected for reasons independent of its observable symptoms, the problem confronting mimics is that by the time observable manifestations of an emotion first play any strategic signaling role, they already have a long evolutionary history. The complex and multi-dimensional links between specific emotions and facial expressions, eye movements, the pitch and timbre of the voice, body language, and a host of other observable details were well-entrenched long before those observable markers could have begun to function as strategic signals. This is problematic because complex signals are more difficult to mimic than simple ones.

Recall that in the case of a dung beetle, a small initial mutation was unlikely to fool even the most nearsighted predator unless the beetle already happened by chance to look almost like a fragment of dung. Similarly, a small mutation in an individual in whom sympathy is not already present is unlikely to make that person more likely to be mistaken for a sympathetic person unless he already happened, purely by accident, to appear almost sympathetic.

Although the brain is a marvelously flexible organ, nothing in our theoretical understanding of how it functions suggests that it ought to be able to summon the complex suite of observable manifestations normally associated with an emotion when the emotion itself is not present. Skilled actors are prized for their ability to convey the emotions called for by their scripts. But they must rehearse to perform convincingly. In everyday interactions, rehearsal is difficult, because emotional cues often arise unpredictably in real time.

³ Similar reasoning applies in the case of commitment problems that entail deterrence. It will often be prudent to exact revenge against an aggressor, even at considerable personal cost, when doing so would help create a reputation that will deter future aggression. Self-interested rational persons with perfect self-control would always seek revenge whenever the future reputational gains outweighed the current costs of taking action. As before, however, the gains from a tough reputation come only in the future while the costs of vengeance-seeking come now. A person may know full well that it pays to be tough, yet still be tempted to avoid the current costs of a tough response. Thus an angry person may be more likely to behave prudently than a merely prudent person who feels no anger. See Frank 1988, chapter 3.

In any event, the claim that emotions are accompanied by characteristic observable signals is not in dispute. As Charles Darwin observed and as has since been confirmed by numerous other investigators, emotional states within the brain produce characteristic suites of autonomous nervous system responses that are visible to external observers (see, for example, Darwin 1872; Ekman et al. 1980; Ekman/Rosenberg (eds.) 1997; Fernandez-Dols et al. 1997). There is also evidence that people rely on such cues when they interact with others in social dilemmas (see Sally 2000). As Fehr and Fischbacher point out, some studies suggest that people are often unable to interpret these cues in brief interactions with strangers. But that does not mean that we cannot make accurate assessments about the emotional makeup of people we know well.

I do not claim that character judgments are always well-founded. Even close friends sometimes disappoint. Indeed, I argued in my earlier work that even if we grant the existence of reliable signals of emotional commitment, the resulting equilibrium must entail a mixed population of cooperators and defectors (see Frank 1987). In any population consisting only of cooperators, no one would be vigilant, and opportunities would thus abound for defectors. In a mixed population, cooperators can survive only by being sufficiently vigilant and skilled in their efforts to avoid good mimics. In short, mimicry is important. But it does not always preclude strategic signaling.

5. Concluding Remarks

In my variant of the green beard hypothesis, people resolve social dilemmas by virtue of their ability to identify and interact selectively with trustworthy partners. Fehr and Fischbacher object to my argument on both empirical and theoretical grounds. The core of their empirical objection is that subjects in laboratory experiments lack the necessary predictive accuracy to render cooperation a winning strategy. Their theoretical objection amounts to a claim that strategic signaling is impossible because of mimicry. Neither objection withstands scrutiny.

My hypothesis does not depend on people being able to identify trustworthy trading partners on the basis of limited interaction with randomly assigned strangers in the laboratory. What is necessary is that they be able to make sufficiently accurate predictions about the trustworthiness of partners they are free to select. Most people feel confident that they can make such predictions about at least some persons they know well. The experiments performed by Fehr and Fischbacher provide no reason to question that belief.

The theoretical objection raised by Fehr and Fischbacher is similarly unpersuasive. They argue that if there were an observable signal of trustworthiness, untrustworthy individuals would eventually mimic it, thereby eliminating its capacity to convey information. Perfect, costless mimicry would indeed make strategic signaling impossible. But mimicry always entails cost and delay, and a careful reading of the theoretical literature on animal signaling suggests why signals of emotions might be especially difficult to mimic. In any event, abun-

dant instances in nature confirm that strategic signals often survive the impact of mimicry. And evidence also confirms that the experience of specific emotions alters people's behavior and appearance in statistically reliable way.

In sum, the arguments and evidence offered by Fehr and Fischbacher leave the green beard hypothesis still kicking. If anything, its plausibility is enhanced by their finding that people can predict cooperation at significantly better than chance accuracy levels even on the basis of limited interaction with strangers.

Bibliography

- Ainslie, G. (1992), *Picoeconomics*, New York
- Andreoni, J. (1995), Warm-glow versus Cold-prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments, in: *Quarterly Journal of Economics* 110.1, 1–21
- Axelrod, R. (1984), *The Evolution of Cooperation*, New York
- Becker, G./G. Stigler (1977), De Gustibus Non Est Disputandum, in: *American Economic Review* 67, 76–90
- Bodvarsson, O. B./W. A. Gibson (1994), Gratuities and Customer Appraisal of Service: Evidence from Minnesota Restaurants, in: *Journal of Socioeconomics* 23, 287–302
- Bruell, J. (1970), Heritability of Emotional Behavior, in: *Physiological Correlates of Emotion*, edited by P. Black, New York
- Darwin, C. (1872), *The Expression of Emotions in Man and Animals*, Chicago 1965
- Dawkins, R. (1976), *The Selfish Gene*, New York
- Ekman, P./W. V. Friessen/S. Ancoli (1980), Facial Signs of Emotional Experience, in: *Journal of Personality and Social Psychology* 39, 1125–1134
- Ekman, P./E. Rosenberg (eds.) (1997), *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, New York
- Fernandez-Dols, J./M. Sanchez/F. Carrera/M. Ruiz-Belda (1997), Are Spontaneous Expressions and Emotions Linked? An Experimental Test of Coherence, in: *Journal of Nonverbal Behavior* 23, 163–177
- Frank, R. H. (1987), If *homo economicus* Could Choose his Own Utility Function, Would he Want one with a Conscience? In: *American Economic Review* 77, 593–604
- (1988), *Passions Within Reason*, New York
- (2004), *What Price the Moral High Ground?*, Princeton
- /T. Gilovich/D. Regan (1993), The Evolution of One-Shot Cooperation, in: *Ethology and Sociobiology* 14, 247–256
- Gould, S. J. (1977), *Ever Since Darwin*, New York
- Hirshleifer, J. (1987), On the Emotions as Guarantors of Threats and Promises, in: *The Latest on the Best: Essays in Evolution and Optimality*, J. Dupre (ed.), Cambridge, 307–26
- Hume, D. (1740), *A Treatise of Human Nature*, Oxford 1978
- Rapoport, A./A. Chammah (1965), *Prisoner's Dilemma*, Ann Arbor
- Sally, D. (1995), Conversation and Cooperation in Social Dilemmas: A Meta-Analysis of Experiments from 1958 to 1972, in: *Rationality and Society* 7, 58–92

- (2000), A General Theory of Sympathy, Mind-Reading, and Social Interaction, with an Application to the Prisoners' Dilemma in: *Social Science Information* 39.1, 567–634
- Schelling, T. C. (1960), *The Strategy of Conflict*, New York
- Smith, A. (1759), *The Theory of Moral Sentiments*, New York 1966
- Tinbergen, N. (1952), Derived Activities: Their Causation, Biological Significance, and Emancipation During Evolution, in: *Quarterly Review of Biology* 27, 1–32